International Society of the Learning Sciences

Towards Asynchronous Data Science Invention Activities at Scale

Rafael Shalala, Ofra Amir, Ido Roll
rafael.s@campus.technion.ac.il, oamir@technion.ac.il, roll@technion.ac.il
Technion, Israel Institute of Technology

Abstract: Invention activities are carefully designed problem-solving tasks in which learners are asked to invent solutions to unfamiliar problems prior to being taught the canonical solutions. Invention activities are typically used in the classroom setting. As online education becomes increasingly common, there is a need to adapt Invention activities to the asynchronous nature of many courses. We do so in the context of an introductory undergraduate data science course. Using an online programming environment, students work on the tasks in pairs, without instructor support. We analyze the invention process and outcomes from two Invention activities on the challenging topics of classification and clustering. Detailed analysis of recordings of six student pairs shows how activity design supports insights at three levels: nature of models (e.g., the need to normalize); domain concepts (e.g., types of errors), and procedural solutions (e.g., weighting errors). We describe the activities, their design, and their outcomes.

Introduction

In Data Science, methods and procedures are defined and implemented in order to extract information and knowledge from datasets. As students often have very little relevant prior knowledge and experiences in these areas, teaching these methods is challenging (Berman et al., 2018). Data science literacy requires knowledge of statistics, understanding of data, and often fluency in programming. Thus, while students often follow the given procedures, they fail to acquire meaningful understanding of relevant concepts.

To address this challenge, we evaluate the benefits of introducing Invention activities to an introductory data science course. In Invention activities, students are asked to develop naive methods to solve problems prior to being taught an expert solution (Loibl, Roll, & Rummel, 2017; Schwartz & Martin, 2004). Such activities help students acquire meaningful experiences, on which future instruction builds (Schwartz, Sears, & Chang, 2012). Invention activities and other similar approaches were shown to improve students' understanding and provide strong foundations for future learning, mainly in the domain of statistics (Holmes, Day, Park, Bonn, & Roll, 2014; Kapur & Bielaczyc, 2012; Loibl et al., 2017). However, it is unclear whether such an approach would also be effective for learning more complex data-science concepts, especially when requiring programming.

A second challenge that we address in this work is the facilitation of Invention activities asynchronously, without instructor support. As online education becomes increasingly prevalent, in both informal (such as MOOCs) and university settings, there is a growing need to support meaningful, active learning in this context (Hew, 2016; Roll, Russell, & Gašević, 2018). To this end, we design activities that are facilitated remotely and asynchronously, via Zoom, without teaching staff support.

We present a case study of designing and deploying remote, collaborative, Invention activities that engage students in problem-solving tasks prior to instruction. We focus on the invention process itself and its outcomes, and discuss lessons learned and implications for the design of asynchronous Invention activities at scale.

Background

In traditional forms of science and math instruction, teachers explain core concepts, and then ask students to apply them in practice problems. Problem-solving followed by instruction (PS-I) flips the traditional approach by first engaging learners in problem solving before the teacher explains the related concepts (Loibl et al., 2017).

Invention activities are a class of the PS-I approach. These are carefully designed problem-solving tasks (Schwartz & Martin, 2004) in which learners are asked to invent general solutions for the given problems. This process helps learners acquire an intuitive understanding of the main domain concepts prior to being taught expert solutions through instruction (Loibl et al., 2017). It is done through the use of contrasting cases which highlight specific features of the domain (Schwartz, Chase, Oppezzo, & Chin, 2011). Invention-based approaches have been shown to boost conceptual learning and transfer to novel situations (Kapur, 2016; Schwartz & Bransford, 1998; Schwartz et al., 2011; Schwartz & Martin, 2004).

Building on these successes, two main challenges motivate the current work. First, the effectiveness of PS-I approaches depends on the type of knowledge being taught (Chase & Klahr, 2017), and its applicability to data science education has yet to be evaluated. Data science is intrinsically complex, as it combines statistics, programming, and big data. Each of these topics is new and challenging for students (Berman et al., 2018). Thus, there is a concern that their combination is too cognitively demanding for engaging in a productive invention process. Second, Invention activities are typically used in classroom settings. Thus, the teacher is often available to support students in their learning (Kapur & Bielaczyc, 2012). Furthermore, students are likely to stay on task even when facing challenges. However, given the current global pandemic, and to support adoption at-scale, we sought to implement Invention activities asynchronously, as homework assignments, without teacher support.

We designed two Invention activities in which students were asked to invent and implement quantitative methods for evaluating the quality of classifiers and for evaluating the quality of clustering methods. We collected recordings of several students who worked on these activities in pairs, and analyzed them. Using this data, we tried to answer our main research question: how to design asynchronous Invention activities to support data science learning for undergraduate level students?

The main contributions of this work are twofold: (1) providing a design approach and rationale for asynchronous Invention activities that could support their adoption at scale, and (2) demonstrating the efficacy of Invention activities for data science by mapping the outcomes of students' invention process to design features of the activities.

Method

To better understand the outcomes of the invention process, and how these were afforded by our design choices, we focus on analyzing students' invention processes and outcomes while working on the activities.

Procedure and Participants

We ran four Invention activities that were followed by lectures as part of an undergraduate level introductory data science course. The first two activities served as a pilot. The latter two activities covered the topics of classification assessment and clustering assessment. They were written using the Jupyter notebook (Perkel, 2018) web application and used Python as the programming language (Python was used for all programming activities in the course). The Jupyter notebook web application allows users to create and edit documents that contain code, text and visualizations. Students worked on the Invention activities in pairs, at their own time, and from their homes. Students were asked to submit their Jupyter notebooks, including their solutions, one day prior to the lecture. Students received the assignments about a week prior to the lecture and could choose when and for how long to work on the activities.

The lecture instruction began with an overview of the students' solutions, followed by teaching of the expert solution or solutions. The overview of the students' solutions included discussions on the differences and trade-offs between them.

All students in the course were asked to complete the activities and were invited to participate in the study. Those who consented were asked to record themselves while working on the activity (while sharing the screen where they edit their code) and share it with the study team, and were given a compensation of $15. Activities took on average 70 minutes (min: 50, max:87). Six student pairs participated in the study (6 males, 6 females). Two pairs participated in both activities. In total, four pairs participated in each activity. Participants had no prior experience with this teaching approach.

Materials

The activities were delivered using code and text embedded in Jupyter notebooks. Each Invention activity included five consecutive tasks:

1. *Introduction* – Students were given a context story. For example, a story about the need for classification of COVID-19 at-risk population according to their medical information, and the goal of a company to develop such a classifier.
2. *Contrasting cases* - Students were presented with two cases that supported intuitive comparisons. Students were asked to choose between these cases and explain their choice. For example, choosing between two classifiers according to their classification results.
3. *Invent a numeric measure* - Students were asked to create a numeric measure for the presented problem. For example, "Suggest a numeric measure to estimate the quality of a classifier, higher value indicates a better classifier".

4. *Implement the suggested measure* - Students were asked to implement their suggested measure. For example, "Implement your suggested measure by completing the following methods that get as input the classifier results and the real data".
5. *Test and reflect* - Students were instructed to test the measure using the examples given in the contrasting cases and reflect upon the outcomes. For example, "Use your suggested measure to examine the classifiers presented in task 2. Do the results support your choice?".

Classification Assessment Invention Activity

The main goal of this activity was to deliver two core concepts in classification: (1) accuracy score is not sufficient for evaluating a classifier and might be misleading, and (2) recognizing the significance of the different types of classification mistakes, namely false positive (wrongly classifying a negative case as positive) and false negative (wrongly classifying a positive case as negative). The introduction described the need for a classifier to identify COVID-19 at-risk populations based on medical data (see Figure 1). The goal of this story was to get students to think about the quality of classifiers in the context of a real-world example. Next, students were asked to choose between two classifiers (contrasting cases) that were tested on data of 40 people, of which only two people were at-risk. One classifier was more accurate, but failed to classify the at-risk people correctly, while the other, though less accurate overall, succeeded in classifying one of the at-risk people correctly. These contrasting cases aimed to highlight the tension between a classifier's accuracy and its ability to avoid critical mistakes.

The next task was to suggest a measure for the quality of classifiers and implement it. Students were provided with code that computes the basic accuracy score of a classifier, i.e., the percent of instances that were correctly classified. They were asked to fill in new methods that propose other measures for the classifiers' quality. The purpose of providing code for basic accuracy was twofold: first, driving students to think of alternative, more elaborate, solutions. Second, basic code that could be edited, scaffolded the process and reduced the risk of time-consuming programming bugs, allowing students to focus on the conceptual challenge of the activity. We further provided students with the name of the method ("classifier_measure1") and its signature which specified the input for the method - two arrays, one for the predictions made by the classifier and one for the ground-truth classification of the test instances. Finally, students were asked to test their implemented measure on the classifiers and reflect upon the choice they have made when choosing their preferred classifier, as well as on the measure they invented.

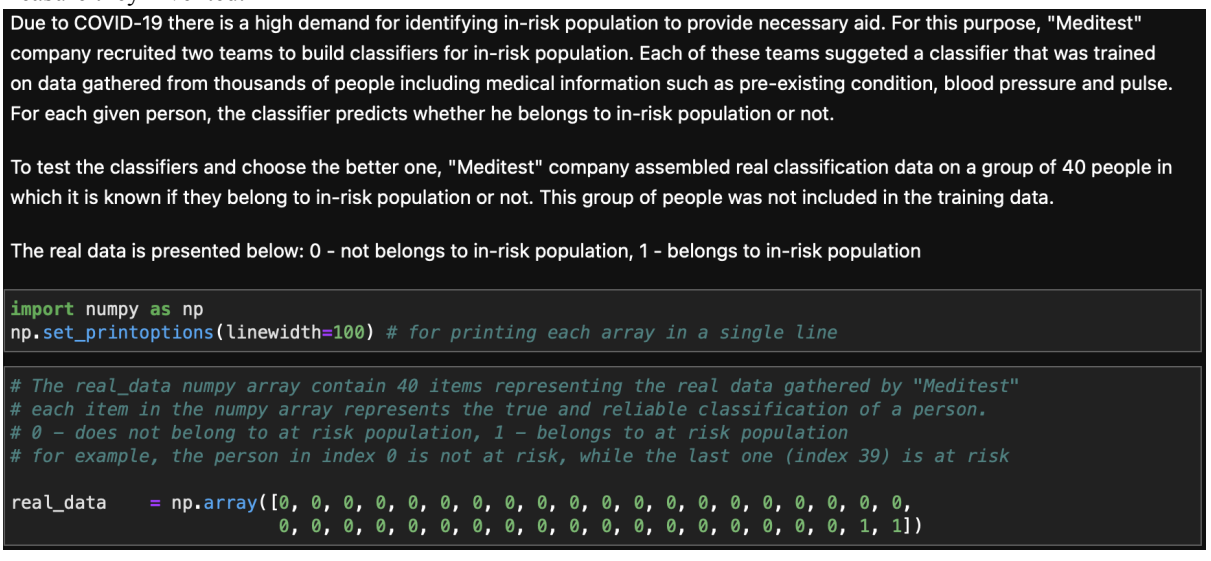

Figure 1. The Classification Assessment activity introduction story

Clustering Assessment Invention Activity

The main goal of this activity was to help students develop an intuition for how to assess a clustering method and give an example of the utility of clustering. The introductory story described an attempt to help students choose academic courses by presenting information about the interest level and difficulty of the courses, and the intent of the students to divide the data into three groups. Next, students were asked to choose between two clustering methods (Figure 2). The clustering on the left provides better separation between the groups, as there is less overlap between groups B and C. Similar to the Classification Assessment activity, the students' next task

was to suggest and implement a measure for the goodness of a proposed division of data points to clusters, and finally, test the implemented measure on the provided clustering methods and reflect upon the choice they have made when choosing their preferred clustering method. In contrast to the classification activity, this activity addressed an unsupervised learning setting, where there is no available ground truth categorization. To support students' coding, we provided them with two auxiliary methods - one that extracts all points belonging to a particular cluster, and one that computes the distance between two data points.

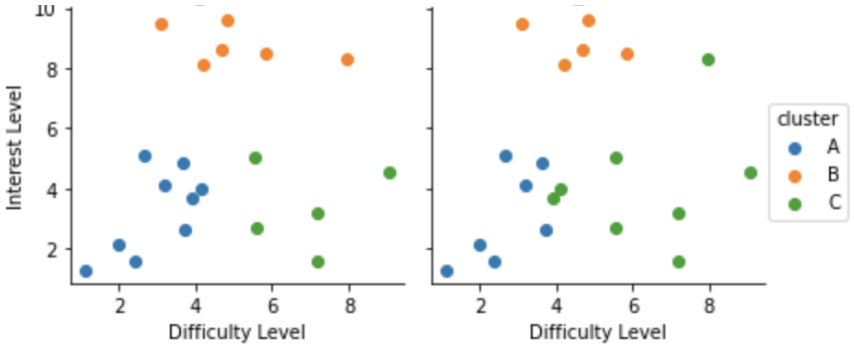

Figure 2. The Clustering Assessment activity choice between two clustering methods

Results

We describe students' outcomes and processes as identified from their video recordings.

Observed Outcomes

The three authors analyzed video recordings from students' invention processes. Each invention process was segmented by turn-taking between the partners. Each segment then received a label (e.g., "identifying the importance of false negatives", "normalizing the measure according to sample size"), and these were clustered into themes. Pretty early it became clear that while students had different interactive patterns and participation models, they reached a finite set of outcomes. We repeated the procedure for two invention processes until we had reached saturation and no new themes were identified. Overall, three categories of students' outcomes were identified:

1. *Conceptual Insights* – insights that are related to the core domain, such as the distinction between different types of classification mistakes (e.g., false negatives vs. false positives).
2. *Design Approaches* - the students' approaches to formalizing their suggested measures, such as using the average between clusters centers to assess the quality of a clustering method.
3. *Nature of Model Insights* - insights that are related to the design of a quantitative measure that are not specifically related to the domain itself, such as considering a measure's boundaries.

Students' outcomes from the two activities are summarized in Table 1 and Table 2. The outcomes in each category are presented alongside examples from transcripts of the activities.

Classification Assessment Activity Outcomes

All four pairs reached the conceptual insights presented in Table 1. That is, all pairs noticed that accuracy is not sufficient for assessing the quality of classification, and also noticed that misclassifying an at-risk person as not at-risk (false negative) is more critical than classifying a person who is not at-risk as at-risk (false positive). The design approaches were evenly distributed between the pairs. Two pairs chose to integrate the accuracy rate and the false negatives rate into a single measure. For example, one pair decided to reduce the false negatives rate from the accuracy score. The other two pairs decided to assign weights according to the type of mistakes that were made. For example, one pair assigned a higher weight (0.6) for false negatives and a lower weight (0.4) for false positives, so the total score was affected more by false negatives. Regarding the nature of model insights, all pairs reached the insight of the importance of providing a general solution that applies to different sample sizes by using normalization in their suggested measure. Two pairs that chose to integrate accuracy and false negative rate paid attention to a case in which the measure result might be negative and modified the measure such that its boundaries will be between 0 and 1.

Clustering Assessment Activity Outcomes

All four pairs achieved the conceptual insight that a clearer separation between clusters indicates a better clustering method. All four pairs focused their design approaches on within cluster distances statistics such as the average of the within cluster distances averages, or the average of within cluster maximum distances. One of the pairs suggested a second measure that focuses on between-cluster distances statistic, and suggested using the average of distances between clusters' centers as a measure for a better clustering method. Another insight that relates to the nature of models was the distinction between choosing average or maximum distance as a statistic. The students realized that when choosing a worst-case approach such as choosing the maximum distance, a single extreme case determines the score for the entire cluster.

Table 1: Observed students' outcomes gained from the Classification Assessment Invention activity. The (PX) mark indicates which of the participating pairs is transcribed

Category	Outcomes	Transcripts
Conceptual Insights	Classification accuracy is not enough	"The second classifier is more accurate, but I don't think we should look at it that way. My opinion stays the same, I still prefer the first one, what do you think?" "I think we should do as we said earlier and consider the more critical mistakes" (P2)
	False negatives vs. false positives	"The question is what is more critical, classifying as at-risk while actually not at-risk, or vice versa?" "Let's think about it, the goal is to identify if you are at-risk. Basically, thinking that you are at-risk while you are not is less dangerous" (P1)
Design Approaches	Integrate accuracy score with critical mistakes	"Look, we can think of something that gives additional weight to specific mistake, but eventually we don't want to neglect the accuracy score, because having many mistakes, even if they are not critical, is not good either" "Maybe we should combine critical mistakes and accuracy somehow" (P1)
	Assign higher weights for critical mistakes	"Maybe we should just give a higher weight for more critical mistakes. I mean, maybe we'll give a 3/5 weight when missing at-risk person and 2/5 for missing not at-risk person" (P2)
Nature of Model Insights	Normalization	"There is something that bothers me, that this measure will be good only for a test group at the same size as in our case, but if I want a more general measure disregarding the test set size, it won't work as we want" "So, let's divide it on the size of the test set to normalize it" (P4)
	Boundaries	"Wait. what will we do if the false negative rate is higher than the success rate, it will lead to a negative result, no?" (P3)

Table 2: Observed students' outcomes gained from the Clustering Assessment Invention activity. The (PX) mark indicates which of the participating pairs is transcribed

Category	Outcomes	Transcripts
Conceptual Insights	Clear separation indicates better clustering method	"Intuitively I want to choose the first clustering method, since each cluster has its own boundaries and seems more clearly separated" "I agree, in this clustering method you can actually draw a clear separation line between the clusters" (P5)
Design Approaches	Within cluster distance statistic	"For each cluster, the points belonging to it should be closer to each other" "So, you mean that the average distances within each cluster should be lower to indicate a better clustering method" (P2)
	Between-clusters distances statistic	"We can calculate the center of each cluster and examine the distance between the clusters' centers" (P6)
Nature of Model Insights	Worst case vs. average	"If we choose maximum distance in a cluster, then a single case determines for the whole cluster" (P1)

Process Analysis

As described earlier, the Invention activities in this study were composed of five consecutive tasks: read a context story, choose between a pair of contrasting cases, suggest a numeric measure for the given problem, implement the measure, and finally, test the measure on the examples given in the contrasting cases, and reflect upon the initial intuitive choice. Our process analysis focuses on linking between the activity structure and the phases that students went through while engaging with the activity. We break down the students' engagement into three phases: Analysis, Invention and Verification.

1. *Analysis* – in this phase students are introduced to the problem and develop their conceptual insights. In both activities, and for all pairs, the conceptual insights were supported by inviting students to engage with the pair of contrasting cases (task 2).
2. *Invention* - This is the main and longest phase of the activities in which the students' inventions are developed and solutions are designed and implemented. In both activities, and for all pairs, this phase occurred while engaging with tasks 3 and 4 - suggesting a measure and implementing the measure. This phase includes two intertwined components: Ideation and Implementation. Ideation is the process in which students discuss various aspects of their suggested solution and develop its fundamentals. Generally, the discussions refer to the parameters that should be taken into account, how they are formulated, and various nature of model aspects such as the measure boundaries. Implementation is the process in which students write code to implement their ideas. The Implementation phase helped the students to get into low-level details they have not paid attention to in the ideation phase, further refining their solutions. In two of the students' works, there was a clear distinction between these phases, the students developed their final solution before engaging with the implementation task, and only then started to implement it. In the remaining works, students went back and forth between those phases. For example, one pair decided to stop the ideation phase to implement false negatives and false positives counters, tested it, and then returned to develop their solution further.
3. *Verification* – in the final phase students test their suggested measure and reflect upon their work. This phase occurs while engaging with the final task (task 5) in which the students were asked to return to the beginning of the activity and test their implemented measure on the contrasting cases from task 2 and reflect upon their initial choice. This phase highlights the benefit of working in a code-based environment which enables implementation and testing cycles.

We demonstrate the invention process by describing in detail the Classifier Assessment activity of a single pair (P1). We describe the different phases, their duration and outcomes, supported by the activity transcripts.

Analysis (minutes 0-7) – In this phase, the students engaged with tasks 1 and 2. In the first two minutes they read the introduction story (task 1) and in the following five minutes they discussed the contrasting cases, attempting to pick the better classifier (task 2). Through these discussions they gained two conceptual insights: (1) *classification accuracy is not enough,* "The first classifier had three mistakes, the second had two, but the first did successfully classify one at-risk person while the second did not", and (2) *false negatives are worse than false positives*:

> Student 1: "The question is which mistake is more critical, classifying as at risk while actually not at risk, or vice versa?"
>
> Student 2: "Let's think about it, the goal is to identify if you are at risk. Basically, thinking that you are at risk while you are not is less dangerous"

Invention (minutes 7-37) – Based on the conceptual insights gained in the analysis phase, the students next tried to suggest a measure (task 3). In this work, there was a clear distinction between ideation and implementation - the students finalized their solution idea in the ideation phase and implemented it exactly as suggested in the implementation phase. In the *Ideation* part (minutes 7-27), the students discussed the parameters that should be considered and formulated the measure, while also addressing the general nature of model aspects. First, they came up with the idea of integrating the accuracy score with critical mistakes:

> Student 1: "Look, we can think of something that gives additional weight to specific mistakes, but eventually we don't want to neglect the accuracy score, because having lots of mistakes, even if they are not critical, is not good either".
>
> Student 2: "Maybe we should combine the critical mistakes with the accuracy somehow".

Next, they came up with a concrete way in which they can integrate the different parameters, "OK, so let's say we have our accuracy rate, the question is what exactly we do with the false negatives and false positives".

Finally, they raised the issue of the boundaries of the model (nature of models insight):

Student 1: “OK, so the accuracy rate is our upper limit”
Student 2: “Wait, but can it be smaller than zero if it is really bad?”

The *Implementation* (minutes 27-37) step was mostly technical. The student wrote code for their measure, making use of the provided auxiliary method for computing accuracy.

Verification (minutes 37-50) - After implementing the measure, the students moved to the final task of testing the measure using the classifiers presented in the contrasting cases (task 2) and reflected on their work. Interestingly, when *Testing* their measure (minutes 37-42), the students tried to predict the output they expect to get from the measure before running their code, “Let’s verify we know what to expect, to verify that [the code] is correct”. They also went beyond the required testing (of the two provided classifiers) and created a new test-case to further examine their measure by modifying the raw data to include more critical mistakes:

Student 1: “We can add more critical mistakes to the raw data and verify we get a reduced score; you want to try?”
Student 2: “Sure”

Finally, in their *Reflection* on the invented measure (minutes 42-50), the students expressed satisfaction from their work, “I am proud of us, this measure is not bad at all”, while acknowledging the limitations of their measure, “Generally, our measure is not suitable for small data such as we got here”. They further noted that there are likely other solutions:

Student 1: “Surely there are other ways to measure classification besides addressing critical mistakes, but eventually you have to give more importance to the type of mistake, so I think we did well with the given time we had, no?”

Student 2: “I think so…”

Discussion

We used carefully designed Invention activities to improve the teaching and learning experience on the topics of classification and clustering in an introductory Data Science course for undergraduate students. The Invention activities took place a couple days prior to the lectures, in which an overview of the students’ solutions was presented, and expert solutions were taught.

The engagement with the activity has led the students to impressive outcomes, including gaining important conceptual insights of the domain, providing valid and complete design approaches for solutions, implementing the suggested solutions while discussing various design aspects including those related to the nature of models (e.g., measure boundaries), and finally, reflecting and analyzing their work.

Design Approach for Data Science Invention Activities

The activity design guided students through a process that was composed of three main phases: Analysis, Invention and Verification. Each of these phases contributed to the outcomes and insights students achieved.

We found that the stories encouraged students to use intuitive knowledge when analyzing the cases and concise contrasting cases helped students notice deep features of the domain. Notably, the contrasting cases were of small data, compared with the typical data science data, in order to enable sense-making. For example, in the Classification Assessment activity, the context story was used as an example in which it was easier to identify the critical mistake - at-risk person that was classified as not at-risk (false negative). The choice between the contrasting cases was used to highlight the problem. The sole method that students knew was the accuracy score, but in the presented problem it was not enough.

Data science makes heavy use of code-based environments. Code-based environments can be useful since they provide tools for easier exploration and testing. In our study, we found that they supported students in iterative ideation-and-implementation. However, they also add a challenge (and extraneous cognitive load) since coding requires technical skills that are not the main focus of the activity. This might lead students with lower coding skills to frustration. To reduce frustration, we added auxiliary methods that could serve as building blocks for the students’ implementation, such as providing a method that extracts all points in a specific cluster, or a method for calculating the distance between two given points. In the trade-off between open exploration and detailed support, through the use of generic methods (rather than developed answers), we tried not to channel students towards specific solutions.

Asynchronous Invention Activities

Working in an out-of-class online environment opens the opportunity for running large scale Invention activities. On the other hand, it raises various challenges, such as the lack of instructor presence and the total freedom

given to the students. When there is no instructor present to guide the students through the activity, the students might get stuck and disengage since they cannot seek help. The converse is also true - students may search google for answers, and thus get too much support that short-circuit the invention process. In our study, we designed several elements that reduced the risks of dropping out or googling for answers. First, the activities were highly structured, providing clear guidance to students as they work. This helped prevent students from 'getting lost'. Second, we found that having students work in pairs was instrumental. Students consulted with each other, challenged each other, and completed each other's ideas. To reduce the risk of students googling answers, we used general language and did not specifically refer to formal domain terms. For example, we did not name the classification activity 'classifier assessment activity', rather we named it 'Meditest Project'. Another measure we took to prevent these "shortcuts" was to highlight to students in lecture that they do not need to find an optimal solution, and that often there is no such single solution. Instead, they were encouraged to come up with alternative measures.

The study has two main limitations. First, the sample size is fairly small, and the students self-selected to the study. Second, we did not evaluate the learning outcomes beyond the activity itself. Future work will address these limitations by evaluating the efficacy of Invention activities and follow-up instruction with a larger sample.

Conclusion

We put forward a design approach for asynchronous Invention activities for learning challenging concepts in data science. Analysis of the outcomes highlighted key insights that students reached through the invention process. Succeeding in implementing Invention activities at scale can add much-needed interactivity to online education, and specifically to data science education.

References

Berman, F., Rutenbar, R., Hailpern, B., Christensen, H., Davidson, S., Estrin, D., … Szalay, A. S. (2018). Realizing the potential of data science. *Communications of the ACM*, *61*(4), 67–72. https://doi.org/10.1145/3188721

Chase, C. C., & Klahr, D. (2017). Invention Versus Direct Instruction: For Some Content, It's a Tie. *Journal of Science Education and Technology*, *26*(6), 582–596. https://doi.org/10.1007/s10956-017-9700-6

Hew, K. F. (2016). Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.12235

Holmes, N. G., Day, J., Park, A. H. K., Bonn, D. A., & Roll, I. (2014). Making the failure more productive: scaffolding the invention process to improve inquiry behaviors and outcomes in invention activities. *Instructional Science*, *42*(4), 523–538. https://doi.org/10.1007/s11251-013-9300-7

Kapur, M. (2016). Examining Productive Failure, Productive Success, Unproductive Failure, and Unproductive Success in Learning. *Educational Psychologist*. https://doi.org/10.1080/00461520.2016.1155457

Kapur, M., & Bielaczyc, K. (2012). Designing for Productive Failure. *Journal of the Learning Sciences*, *21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Loibl, K., Roll, I., & Rummel, N. (2017). Towards a Theory of When and How Problem Solving Followed by Instruction Supports Learning. *Educational Psychology Review*, *29*(4), 693–715. https://doi.org/10.1007/s10648-016-9379-x

Perkel, J. M. (2018). Why Jupyter is data scientists' computational notebook of choice. *Nature*. https://doi.org/10.1038/d41586-018-07196-1

Roll, I., Russell, D. M., & Gašević, D. (2018). Learning at Scale. *International Journal of Artificial Intelligence in Education*, *28*(4), 471–477. https://doi.org/10.1007/s40593-018-0170-7

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*(4), 475–5223. https://doi.org/10.1207/s1532690xci1604_4

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing Versus Inventing With Contrasting Cases: The Effects of Telling First on Learning and Transfer. *Journal of Educational Psychology*, *103*(4), 759–775. https://doi.org/10.1037/a0025140

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*(2), 129–184. https://doi.org/10.1207/s1532690xci2202_1

Schwartz, D. L., Sears, D., & Chang, J. (2012). Reconsidering prior knowledge. *Thinking with Data*, (650), 319–344. https://doi.org/10.4324/9780203810057